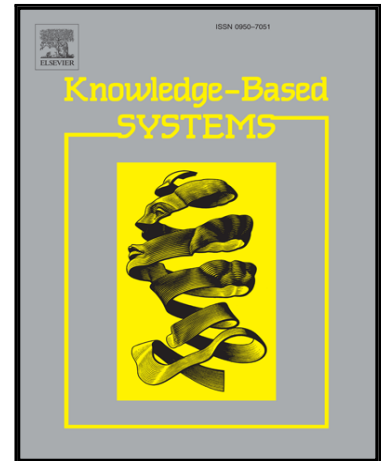


## Accepted Manuscript

Incremental fuzzy cluster ensemble learning based on rough set theory

Jie Hu, Tianrui Li, Chuan Luo, Hamido Fujita, Yan Yang

PII: S0950-7051(17)30294-0  
DOI: [10.1016/j.knosys.2017.06.020](https://doi.org/10.1016/j.knosys.2017.06.020)  
Reference: KNOSYS 3948



To appear in: *Knowledge-Based Systems*

Received date: 21 February 2017  
Revised date: 9 June 2017  
Accepted date: 12 June 2017

Please cite this article as: Jie Hu, Tianrui Li, Chuan Luo, Hamido Fujita, Yan Yang, Incremental fuzzy cluster ensemble learning based on rough set theory, *Knowledge-Based Systems* (2017), doi: [10.1016/j.knosys.2017.06.020](https://doi.org/10.1016/j.knosys.2017.06.020)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Incremental fuzzy cluster ensemble learning based on rough set theory

Jie Hu<sup>a</sup>, Tianrui Li<sup>a\*</sup>, Chuan Luo<sup>b</sup>, Hamido Fujita<sup>c</sup>, Yan Yang<sup>a</sup>

<sup>a</sup>*School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China*

<sup>b</sup>*College of Computer Science, Sichuan University, Chengdu, 610065, China*

<sup>c</sup>*Faculty of Software and Information Science, Iwate Prefectural University, 020-0693, Iwate, Japan*

---

## Abstract

To deal with the uncertainty, vagueness and overlapping distribution within the data sets, a novel incremental fuzzy cluster ensemble method based on rough set theory (IFCERS) is proposed by the idea of combining clustering analysis task with classification techniques. Firstly, on the basis of soft clustering results, the positive region, boundary region and negative region of clustering ensemble are obtained by applying the construction of rough approximation in rough set theory, and then a group structure within data points of positive region is obtained by adopting a fuzzy cluster ensemble method. Secondly, by combining with the supervised ensemble learning method, e.g., random forests, the obtained group structure is used to construct the random forests classifier to classify the data points in boundary region. Finally, all the acquired group structure is used to train the random forests classifier to classify the data points of negative region. Experimental evaluations on UCI machine learning repository datasets verify the effectiveness of the proposed method. It is also shown that the quality of the final solution has a weak correlation with the ensemble size, the parameter setting on the rough approximations construction is appropriate, and the proposed method is robust towards the diversity from hard clustering members.

*Keywords:* Cluster ensemble, Granular computing, Rough sets, Random forests

---

## 1. Introduction

Data clustering is a fundamental unsupervised pattern classification technique that has been used to discover the natural cluster structure of an unlabeled data set [1]. Cluster ensemble is an important extension of the classical cluster analysis technique with the objective to obtain more stable, robust and accurate final result than individual clustering results by aggregating over multiple clustering solutions [2]. At present, cluster ensemble has been widely applied in many domains related to data mining and machine learning, including multimedia data analysis, social network analysis, collaborative filtering, customer segmentation, biological data analysis, outlier analysis, dynamic trend detection, etc [1].

The ubiquitous uncertain, vagueness and overlapping among clusters have posed a great challenge to the cluster analysis task. Granular computing, which has emerged as a new computing paradigm of complex information processing in the area of computational intelligence, provides people with theories, methodologies, techniques and tools to describe and process uncertain, imprecise and incomplete information [3]. Considered as the two major mathematical tools of granular computing, fuzzy set theory (FST) and rough set theory (RST) have been introduced into the clustering method to reveal uncertain structured data set [4–10], which extend the clustering from hard-partition to soft-partition and thus forms an important branch of the data clustering [4, 11]. For example, fuzzy c-means (FCM) and its derivatives, rough k-means and its derivatives, and hybrid rough-fuzzy clustering approaches have been proposed [4]. In fuzzy clustering [12], each pattern is assigned with a partial membership to a given cluster and one object is allowed to assign to more than one cluster at the same time. In rough clustering [6], each cluster is represented with

---

\*Corresponding author.

*Email addresses:* jiehu@swjtu.edu.cn (Jie Hu), trli@swjtu.edu.cn (Tianrui Li), cluo@scu.edu.cn (Chuan Luo), HFujita-799@acm.org (Hamido Fujita), yyang@swjtu.edu.cn (Yan Yang)

Download English Version:

<https://daneshyari.com/en/article/4946130>

Download Persian Version:

<https://daneshyari.com/article/4946130>

[Daneshyari.com](https://daneshyari.com)